

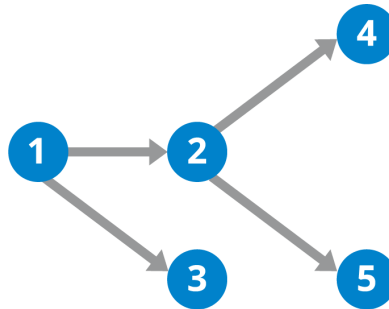


Orquestração de Fluxos de Dados com Apache Airflow

Conceitos Básicos - DAG



O **DAG**, originado da sigla para "Directed Acyclic Graph" (Grafo Dirigido Acíclico em português), é um conceito fundamental no Apache Airflow. Ele representa uma coleção de tarefas organizadas de forma a refletir suas relações e dependências.



Em outras palavras, é um conjunto de vértices (tarefas) e arestas (dependências entre tarefas) onde cada aresta tem uma direção e não há ciclos, o que significa que não é possível voltar a um vértice anterior seguindo as direções das arestas.

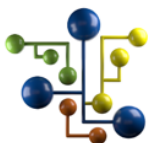
No contexto do Airflow, um DAG não é apenas uma representação teórica, mas também uma definição prática de um fluxo de trabalho.

Cada tarefa em um DAG é uma instância de uma classe de operador, como `PythonOperator` ou `BashOperator`, que define uma ação específica a ser executada.

O Airflow utiliza o DAG para determinar a ordem e o timing das execuções das tarefas, garantindo que as tarefas sejam executadas na ordem correta e apenas quando todas as suas dependências forem atendidas.

O DAG oferece também flexibilidade e robustez na orquestração de fluxos de trabalho. Com ele, os usuários podem definir tarefas complexas, reutilizar componentes, lidar com falhas e reexecuções, e escalar fluxos de trabalho em sistemas distribuídos.

A visualização do DAG na interface web do Airflow também fornece uma representação clara do fluxo de trabalho, facilitando o monitoramento e a depuração.



Equipe DSA

Muito Obrigado!

Continue Trilhando Uma Excelente Jornada de Aprendizagem.